(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 1 096 391 A2

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
02.05.2001 Bulletin 2001/18

(51) Int. Cl.[7]: **G06F 17/30**

(21) Application number: **00118884.6**

(22) Date of filing: **31.08.2000**

(84) Designated Contracting States:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE
Designated Extension States:
AL LT LV MK RO SI

(30) Priority: **26.10.1999 US 428136**

(71) Applicant:
Hewlett-Packard Company,
A Delaware Corporation
Palo Alto, CA 94304 (US)

(72) Inventors:
- **Shmueli, Oded**
  Nofit 36001 (IL)
- **Elad, Michael**
  Haifa 34603 (IL)
- **Greig, Darryl**
  Haifa 34332 (IL)
- **Staelin, Carl**
  Haifa 34973 (IL)

(74) Representative:
Schoppe, Fritz, Dipl.-Ing.
Schoppe, Zimmermann & Stöckeler
Patentanwälte
Postfach 71 08 67
81458 München (DE)

(54) Automatic categorization of documents using document signatures

(57) A method of quickly and automatically comparing a new document to a large number of previously seen documents and identifying the document type. First, provide a plurality of document type distributions, each document type distribution describes layout characteristics of an independent document type and may include a plurality of data points. Each document type distribution includes data derived from at least one basis document signature which may include data defining pixels of a low-resolution image of the independent basis document resolved to between 1 and 75 dots per inch or may include document segmentation data derived from the independent basis document. Next provide a new electronic document. Then create new document signature from the new electronic document. Next, distances between the new document signature and each of the plurality of document type distributions are calculated using an algorithm based on a Bayesian framework for a Gaussian distribution. The distances calculated may be Euclidean distances or may be Mahalanobis distances. Additionally, calculating the distances may include weighting the value given each of a plurality of data points in the document signatures based on the usefulness of each of the plurality of data points in distinguishing between the document signatures. Next, select at least one candidate document type for the new electronic document from among the independent document types described by the plurality of document type distributions. The selection of the at least one candidate document type may include selecting a preselected fixed number of the independent document types or may include selecting the independent document types described by those of the plurality of document type distributions having calculated distances that are within a preselected threshold distance of the smallest of the distances calculated. In addition, the invention provides for a program storage medium readable by computer, tangibly embodying a program of instructions executable by the computer to perform the method steps described above.

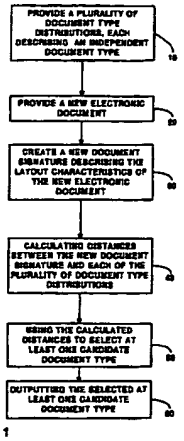


FIG. 1

EP 1 096 391 A2

ument, a low-resolution representation of the document segmentation of the basis document, or some other similar representation of the basis document. The data derived from the at least one basis document signature may include a multiple representative statistic value such as a mean or median value of each of the data values across each of the at least one document signatures.

[0009] The next step is providing a new electronic document. Then a new document signature is created from the new electronic document. The new document signature describes the layout characteristics of the new electronic document and may include data defining pixels of a low-resolution image of the new electronic document, a low-resolution representation of the document segmentation of the new electronic document, or some other similar representation of the new electronic document.

[0010] Next, distances between the new document signature and each of the plurality of document type distributions are calculated. The distances may be calculated using distance measures known in the art, such as Euclidean distance, Mahalanobis distance, an algorithm based on a Bayesian framework for a Gaussian distribution, or other measures. Additionally, distance calculations may weight the value given each of a plurality of data points in the basis document signatures or the document type distributions based on the usefulness of that data point in distinguishing between the various document types or the reliability of that point in specifying a particular document type. The reliability of each of the plurality of data points may be calculated, for example, based on the ratio of the spread of that data point within all basis documents of that document type to a spread of that data point across all of the plurality of the basis documents.

[0011] Based on the distances calculated, at least one candidate document type for the new electronic document is selected from among the independent document types described by the plurality of document type distributions. The selection of the at least one candidate document type may include selecting a preselected fixed number of the independent document types. The preselected fixed number of independent document types may be those described by the preselected fixed number of the plurality of document type distributions calculated to have the preselected fixed number of shortest distances. Alternatively, the selection of the at least one candidate document type may include selecting the independent document types described by those of the plurality of document type distributions having calculated distances that are within a preselected threshold distance of a shortest of the distances calculated. Further, the selection algorithm of the at least one document type may declare that the new electronic document is of a new type.

[0012] In addition, the invention provides for a program storage medium readable by computer, tangibly embodying a program of instructions executable by the computer to perform the method steps described above. Other aspects and advantages of the invention will become apparent from the following detailed description, taken in conjunction with the accompanying drawings and the attached pseudo code listing, illustrating by way of example the principles of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

[0013]

Figure 1 is a flowchart depicting the method of the invention.

Figure 2A is a first sample basis document signature of the low-resolution image type from a first document type.

Figure 2B is a second sample basis document signature of the low-resolution image type from a second document type.

Figure 2C is a third sample basis document signature of the low-resolution image type from a third document type.

Figure 3A is a fourth sample basis document signature of the document segmentation type from the same document type as shown in Figure 2A.

Figure 3B is a fifth sample basis document signature of the document segmentation type from the same document type as shown in Figure 2B.

Figure 3C is a sixth sample basis document signature of the document segmentation type from the same document type as shown in Figure 2C.

Figure 4 is a graph comparing the performance of document segmentation type basis document signatures and low-resolution image type basis document signatures in the method according to the invention.

representative statistics such as mean, median, mode and standard deviations derived from the data in each of the basis document signatures of the independent document type; 5) statistical information derived from the data of a sampling of the basis document signatures of the independent document type; and 6) any combination of the above.

[0022] A first type of basis document signature which may be used is the low-resolution image type which is also known as a "thumbnail" image type. Three examples of low-resolution image type basis document signatures 101, 102, 103 are depicted in Figures 2A - 2C, respectively. The low-resolution image type of basis document signature is achieved by down-scaling the original document image for each basis document of the particular document type. The original document images typically have a resolution of 300 dots per linear inch (dpi). Each of the dots is usually referred to as a "pixel". With an 8 inch by 11 inch document, this corresponds to 2400 pixels by 3300 pixels for a total of 7,920,000 pixels per document. By reducing the resolution of the image to between 3 dpi and 9 dpi an image of between 24 pixels by 33 pixels and 72 pixels by 99 pixels, respectively, is created. These correspond to low-resolution document images with between 729 pixels per document and 7128 pixels per document, or a decrease in the total number of pixels by a factor of between 100 and 1,000. The example low-resolution image type basis document signatures 101 - 103 are at a resolution of 9 dpi and a sample pixel 110 - 112 is indicated on each document signature, respectively. For purposes of this description, a low-resolution document image may be as high as 75 dpi, but is preferably below 15 dpi.

[0023] Often, thumbnail images of document images are created automatically by commercially available document scanning software so that documents can be easily previewed and selected by users. Thus, the thumbnail images that form the document signatures can often be provided with little or no additional computational cost which is important, particularly when processing a large set of documents. It is also possible to use lower resolution images down to 1 dpi or below to further reduce computational and memory requirements, however reducing the resolution below 3 dpi can substantially reduce the accuracy of the method according to the invention as will be discussed below.

[0024] The "thumbnail" images from each basis document of a particular document type are then used to create a document type distribution for that document type using any of the techniques described above. For example, one way to create a document type distribution would be to combine each of the low-resolution type basis document signatures into a single "thumbnail" image that is a "mean image" representing the document type. The method for creating this "mean image" this will depend on whether the thumbnail images from the basis documents are binary or grayscale. Binary pixels are either black or white, while grayscale pixels are defined as a point along a scale between completely black and completely white. Typically a grayscale pixel will be broken into 256 increments, or levels of gray.

[0025] If the thumbnail images are binary, then each pixel is compared to the corresponding pixel on the other basis document thumbnail images. If there are more black pixels than white pixels, the corresponding pixel is set to black in the document type distribution. Similarly, if there are more white pixels than black pixels for a particular pixel location on each of the basis documents, then the corresponding pixel in the document type distribution is set to white. If an equal number of black pixels and white pixels exist for a particular pixel location on each of the basis documents, then the corresponding pixel in the document type distribution is set randomly to black or white.

[0026] If the thumbnail images are grayscale, then each pixel is compared to the corresponding pixel on the other basis document thumbnail images and an average level of gray is calculated. Thus, if there are three basis document thumbnail images and the first pixel of each has a gray level of 25, 175, and 250, respectively, then the corresponding pixel 110 in the document type distribution is set to a level of 150 = (25 + 175 + 250) / 3.

[0027] A second type of document signature that may be used is a document segmentation type. Three examples of document segmentation type document signatures, 104, 105, and 106 are depicted in Figures 3A - 3C, respectively. The document segmentation type of document signature is stylized representation of the document type built from the output of a page decomposition algorithm from each of the basis documents of that document type. Page decomposition algorithms are known in the art and are typically included in commercially available document scanning software. Traditionally the output of a page decomposition algorithm is a collection of geometric shapes marking discrete blocks on the page. The page decomposition algorithms can either provide binary block data or weighted block data dependent on, for example, the font size in a text block, or some other pixel density measure in general. In some cases, the output of the page decomposition algorithm for each basis document can be obtained at no or low computational cost, by simply siphoning the necessary numbers into a file as part of the page decomposition done prior to optical character recognition (OCR) processing of a document.

[0028] The output of the page decomposition algorithm from each of the basis documents are used to create the document segmentation type of basis document signature for that basis document. The basis document signatures for each of the independent basis documents of a particular document type may be combined into a document type distribution using any of the techniques described above. For example, document type distributions may be formed by averaging document segmentation signature data values to create a "mean segmentation image." The averaging process will depend on whether the output of the page decomposition algorithm was binary or weighted. For binary output, blocks with no data are defined by a 0 value while block containing data (text or otherwise) are defined with a 1 value. Each location in a basis document is compared against the corresponding location in the other basis documents. The locations will typically correspond to the pixel locations in the low-resolution image type document signatures. If there

other hand, a method which can reliably determine a small subset of classes containing the correct class in at most some number log(N) guesses in the number of classes, but does not require extensive re-computation upon the addition of a new class, is preferable. One way this can be accomplished is by calculating the distances using an algorithm based on a Bayesian framework for a Gaussian distribution

[0035] If the method of automatically classifying documents according to the invention is to be a preprocessing stage for some "heavier" system of extracting data from documents, the method according to the invention should be able to select between the candidate classes offered to it, and if necessary reject them all. One effective way to accomplish this is to utilize approaches which emerge from the Bayesian decision rule. For purposes of this description, the plurality of document signatures will be denoted by X_k^j. The document class (type) number is represented by $k = 1, 2, 3, \ldots C$ where C is a constant representing the total number of the plurality of document types. The basis document number is represented by $j = 1, 2, \ldots, N_k$ where N_k is the total number of basis documents represented by the k-th document type distribution. We assume that

$$\left\{X_k^j\right\}_{j=1}^{N_k}$$

are drawn from a Gaussian multivariate distribution $G\{M_k, \Sigma_k\}$ where M_k is the multivariate mean and Σ_k is the covariance matrix. Thus, the classification of the new document signature z is done by computing the Mahalanobis distances:

$$d_k = D(z, M_k, \Sigma_k) = (z - M_k)^T \Sigma_k^{-1} (z - M_k) \quad (1)$$

where T is the matrix transpose, and mapping z to the class k_0 with the minimal distance

$$d_{k_0} = min\{d_1, d_2, \ldots, d_C\}.$$

Additionally, calculating distances may include heuristic methods for approximating the covariance matrix of each document type distribution. For clarity, the notation

$$\text{“}\Sigma_k^{-1}\text{”}$$

in equation (1), above indicates the inversion of the covariance matrix rather than a summation.

[0036] Next, based on the distances calculated, at least one candidate document type for the new electronic document from among the independent document types described by the plurality of document type distributions is selected (block 50). For purposes of this description, selecting at least one candidate document type may include indicating that none of the document types described by the plurality of document type distributions are good candidates. If a preselected fixed number of output candidates document types are desired, we may simply choose the preselected fixed number of candidate document types corresponding those of the plurality of document type distributions with the smallest distances. Another option is to choose all the candidate document types corresponding to those of the plurality of document type distributions having a distance within some fixed distance of the minimal distance. For purposes of this description, this second technique will be called "adaptive candidate selection." Clearly adaptive candidate selection will result in a variable number of candidate document types being proposed by the method according to the invention, however the percentage threshold may be adjusted to specify the average number of candidates returned in repeated uses of the method according to the invention. It has been found experimentally that the variance in the number of output candidate document types proposed is low. Thus, the probability is low of the method according to the invention returning an unacceptably high number of candidate types with this technique allowing a variable number of candidates document types to be proposed.

[0037] Figure 6 is a graph which indicates the experimentally derived relative performance of selecting a fixed number of candidate document types and adaptive candidate selection. In the experiment 18 different document type distributions were tested. Each document type distribution was prepared from between 20 and 200 basis document signatures of the low-resolution type at a resolution of 5 dpi. The x-axis of the graph indicates the number of candidate document types that the method according to the invention was allowed to pick as either a preselected fixed number of selections or an average number for adaptive candidate selection. The y-axis of the graph indicates the accuracy of the method according to the invention in percent. The mean performance of the adaptive candidate selection is shown by

$$\Delta_{k_0}(i) = \sqrt{1 - \frac{\text{Spread within Class}}{\text{Spread for all Classes}}} =$$

$$\sqrt{1 - \frac{\frac{1}{N_{k_0}} \sum_{j=1}^{N_{k_0}} \left(X_{k_0}^{j}(i) - M_{k_0}(i) \right)^2}{\frac{1}{\sum_{k=1}^{C} N_k} \sum_{k=1}^{C} \sum_{j=1}^{N_k} \left(X_k^{j}(i) - M_{k_0}(i) \right)^2}} \qquad (6)$$

[0043] Figure 7 is a graph which indicates the experimentally derived relative performance of using a Simple Bayesian method and using a Weighted Bayesian method. In the experiment 18 different document type distributions were tested. Each document type distribution was prepared from between 20 and 200 basis document signatures of the low-resolution type at a resolution of 5 dpi. The x-axis of the graph indicates the number of candidate document types that the method according to the invention was allowed to pick using adaptive candidate selection. The y-axis of the graph indicates the accuracy of the method according to the invention in percent. The mean performance of the Weighted Bayesian method is shown by the solid line 125 while the mean performance of the Simple Bayesian method is shown by the dashed line 126. The results clearly indicate that the Weighted Bayesian method has a marked advantage in accuracy over the Simple Bayesian method when less than four candidates are selected.

[0044] No matter which signature type, resolution, candidate selection technique, or calculation method is chosen, the results of the method according to the invention may be output (block 60) directly to a user, or to an expert system for further processing of the new electronic document.

[0045] In addition to the method described above, another preferred embodiment of the invention is a program storage medium readable by computer, tangibly embodying a program of instructions executable by the computer to perform the method steps described above. In this embodiment, the various steps described above are performed by a computer. In light of this fact and in order to provide a more detailed description of the method according to the invention, a listing of pseudo code for running the method on a computer is attached.

[0046] Although a specific embodiment of the invention has been described and illustrated, the invention is not to be limited to the specific forms or arrangements of parts so described and illustrated. The invention is limited only by the claims.

```
//////
// scoring new document:
//      given the mean and weights for each document type, find the distances of the
//      new document signature z from all the document type distributions
/////
{
    // first need the signature of the new document

    Create signature z of new document;

    // now compute the distance of this signature from all the document type
    // distributions (requires the means and weights for each document type)

    for ( each document type k ) {
        Compute distance d_k from z to document type k ;      (eq (5))
    }

    // finish with a vector of distances d_k for each document type
}

//////
// select candidate document types:
//      given the vector of distances d_k for each document type, return candidate
//      document types, two selection methods
//////

SelectThreshold(t)
// method [1] - t is a value between 0 and 1 such that 100t  is the percentage threshold
{
    // first find the candidate giving the minimal distance and add it to the candidate
```

new electronic document from among the independent document types described by the plurality of document type distributions.

2. The method of claim 1, in which calculating the distances (40) in step (d) includes using an algorithm based on a Bayesian framework for a Gaussian distribution.

3. The method of claim 1, in which:

the at least one basis document signature (101 - 103) in step (a) includes data defining pixels (110 - 112) of a low-resolution image of the independent basis document; and
the new document signature in step (c) includes data defining pixels of a low-resolution image of the new electronic document.

4. The method of claim 3, in which the data derived from at least one basis document signature in step (a) includes a multiple representative statistic value across each of the at least one basis document signatures of each of the pixels of the low-resolution image.

5. The method of claim 3, in which:

the low-resolution image of the independent basis document is resolved to between 1 and 75 dots per inch; and

the low-resolution image of the new electronic document is resolved to between 1 and 75 dots per inch.

6. The method of claim 1, in which:

the at least one basis document signature in step (a) includes document segmentation data (113 - 115) derived from the independent basis document of the independent document type; and

the new document signature in step (c) includes document segmentation data derived from the new electronic document.

7. The method of claim 6, in which the data derived from at least one basis document signature in step (a) includes a multiple representative statistic across each of the at least one basis document signature of document segmentation data.

8. The method of claim 2, in which selecting the at least one candidate document type (50) in step (e) includes selecting a preselected fixed number of independent document types described by the preselected fixed number of the plurality of document type distributions calculated (40) in step (d) to have the preselected fixed number of shortest distances.

9. The method of claim 2, in which selecting the at least one candidate document type in step (e) includes selecting the independent document types described by those of the plurality of document type distributions having distances calculated in step (d) within a preselected threshold distance of a minimal distance calculated in step (d).

10. The method of claim 2, in which the distances calculated in step (d) are Euclidean distances.

11. The method of claim 2, in which the distances calculated in step (d) are Mahalanobis distances.

12. The method of claim 2, in which:

each of the plurality of document type distributions provided in step (a) includes a plurality of data points; and

calculating distances in step (d) includes weighting the value given each of the plurality of data points based on a calculated reliability of each of the plurality of data points.

13. The method of claim 11, in which the calculated reliability of each of the plurality of data points includes the ratio of:

each of the plurality of document type distributions provided in method step (a) includes a plurality of data points; and

calculating distances in method step (d) includes weighting the value given each of the plurality of data points based on a calculated reliability of each of the plurality of data points.

22. A program storage medium of claim 21, in which in which the calculated reliability of each of the plurality of data points includes the ratio of:

a spread of each of the plurality of data points within each of the plurality of document type distributions, respectively, to

a spread of each of the plurality of data points across all of the plurality of document type distributions, respectively.

FIG. 2A

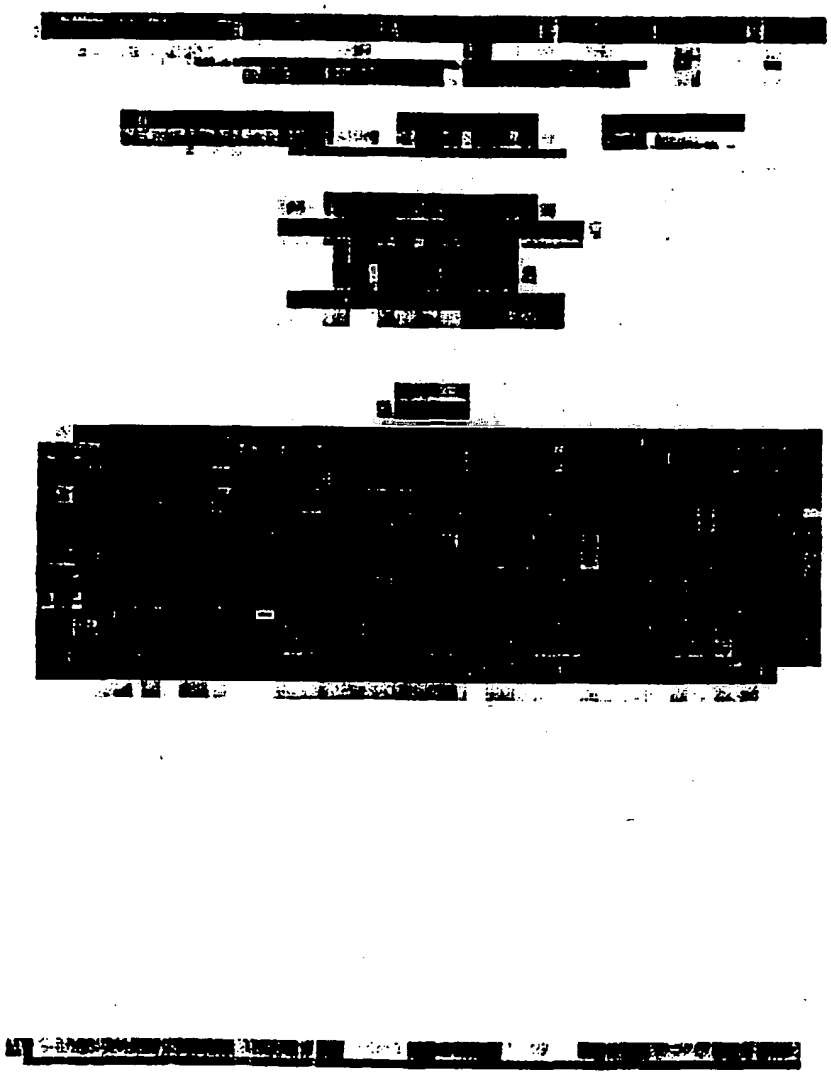


FIG. 2C

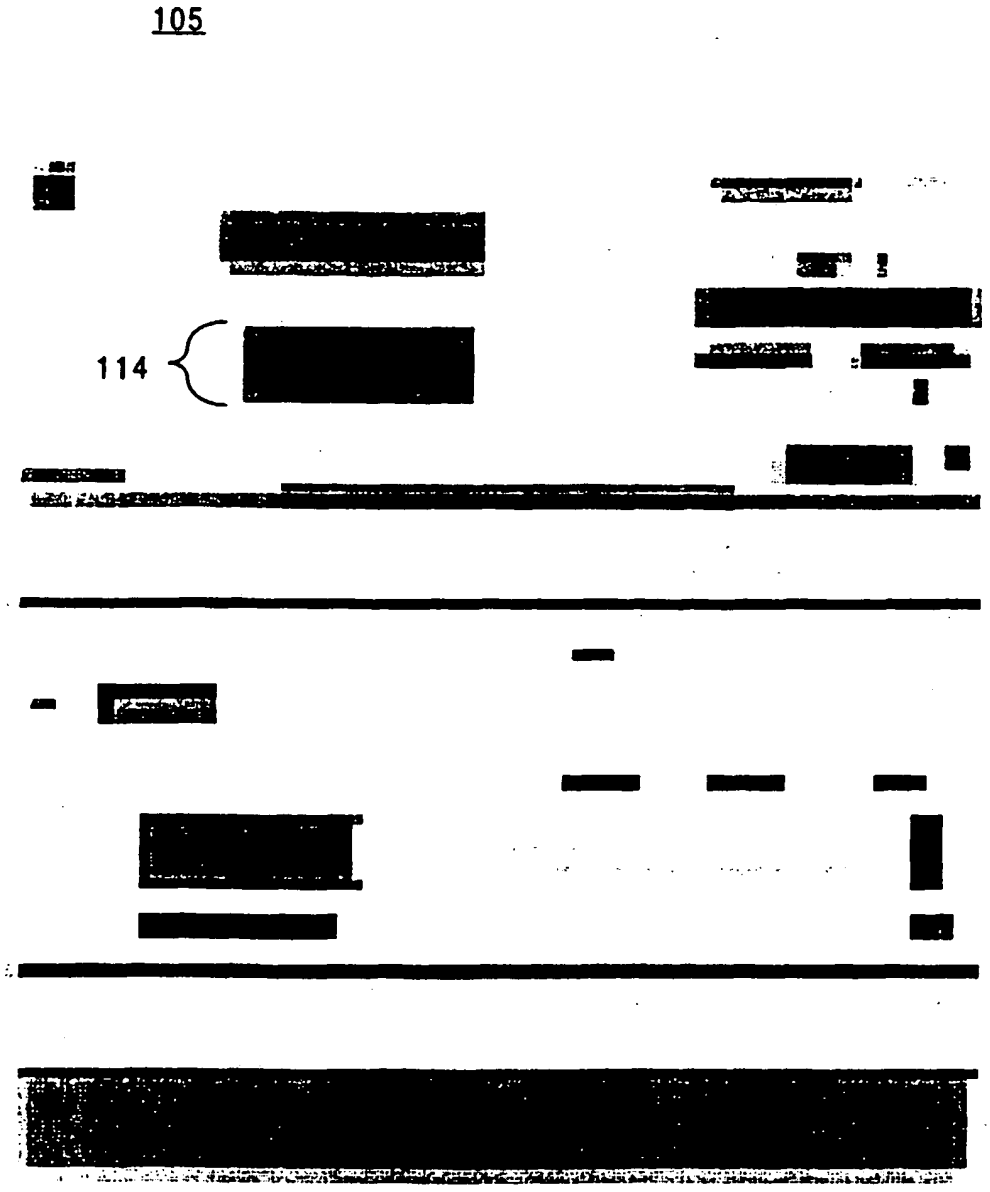


FIG. 3B

FIG. 4

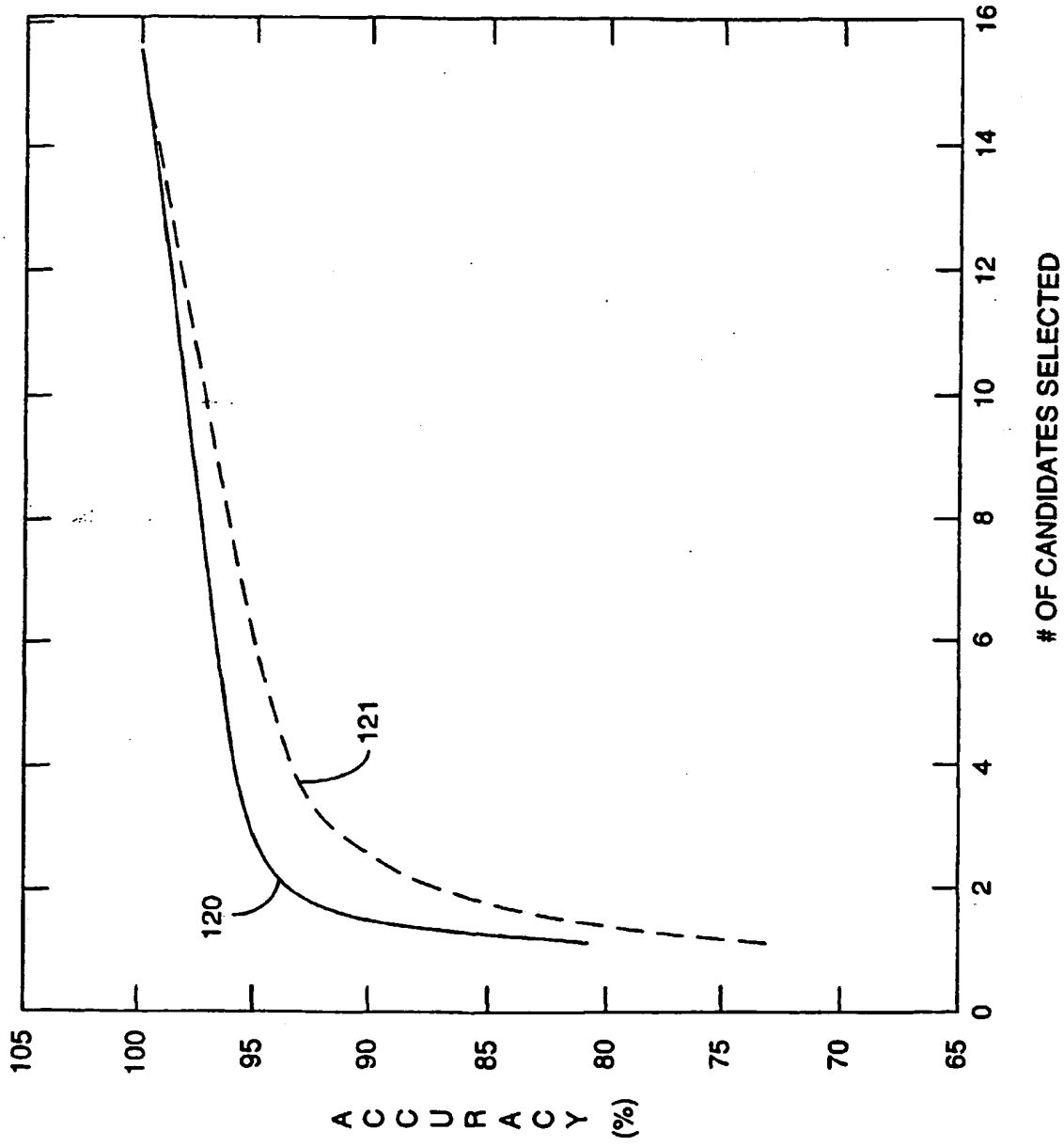
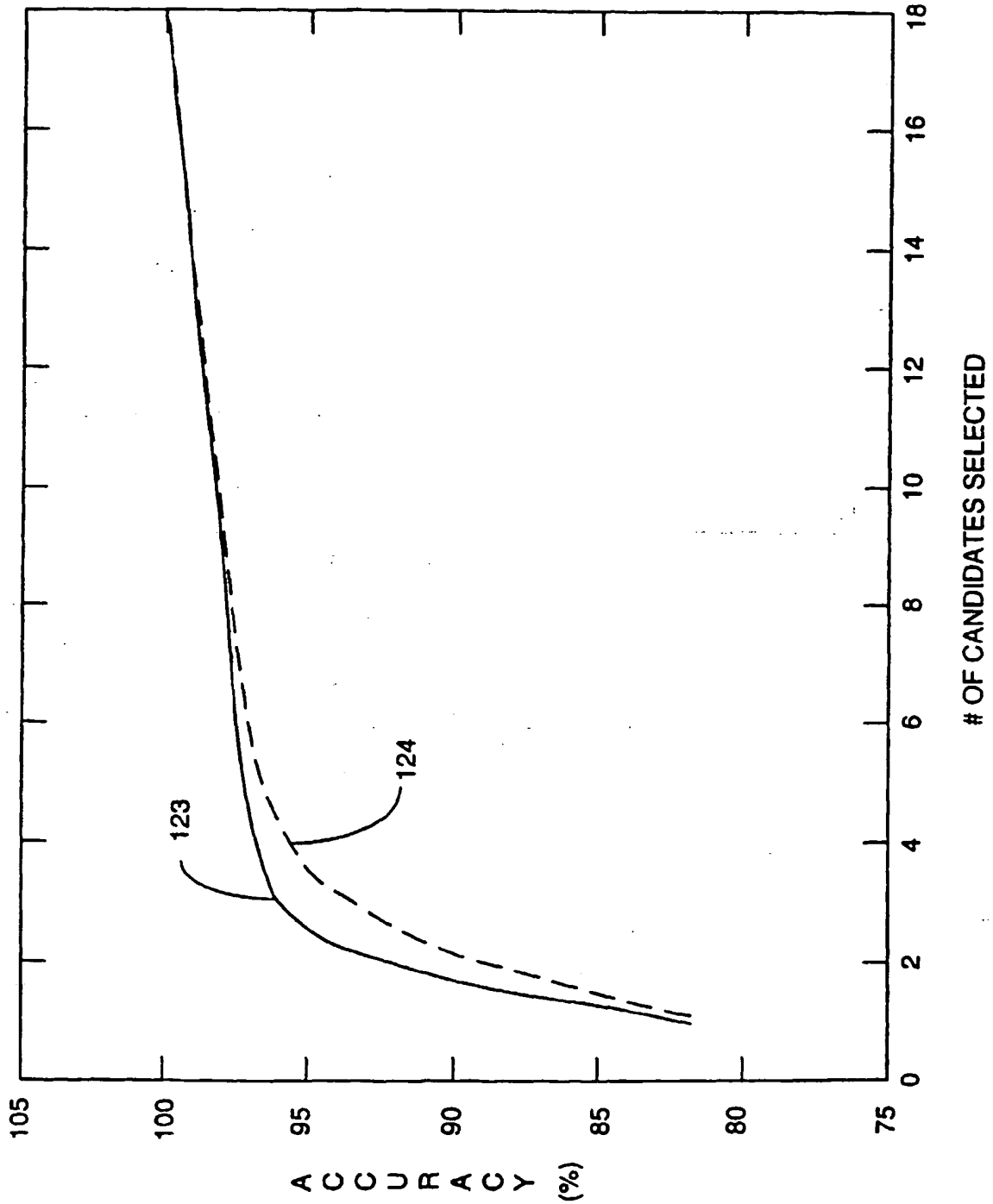

105
100
95
90
85
80
75
70
65
ACCURACY (%)
0
2
4
6
8
10
12
14
16
OF CANDIDATES SELECTED
120
121

FIG. 6


105
100
95
90
85
80
75
ACCURACY (%)
0
2
4
6
8
10
12
14
16
18
OF CANDIDATES SELECTED
123
124

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 1 096 391 A3

(12) EUROPEAN PATENT APPLICATION

(88) Date of publication A3:
26.05.2004 Bulletin 2004/22

(43) Date of publication A2:
02.05.2001 Bulletin 2001/18

(21) Application number: **00118884.6**

(22) Date of filing: **31.08.2000**

(51) Int Cl.7: **G06F 17/30**

(84) Designated Contracting States:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE
Designated Extension States:
AL LT LV MK RO SI

(30) Priority: **26.10.1999 US 428136**

(71) Applicant: **Hewlett-Packard Company, A Delaware Corporation**
Palo Alto, CA 94304 (US)

(72) Inventors:
- **Shmueli, Oded**
  Nofit 36001 (IL)
- **Elad, Michael**
  Haifa 34603 (IL)
- **Greig, Darryl**
  Haifa 34332 (IL)
- **Staelin, Carl**
  Haifa 34973 (IL)

(74) Representative: **Schoppe, Fritz, Dipl.-Ing.**
Schoppe, Zimmermann, Stöckeler & Zinkler
Patentanwälte
Postfach 246
82043 Pullach bei München (DE)

(54) Automatic categorization of documents using document signatures

(57) A method of quickly and automatically comparing a new document to a large number of previously seen documents and identifying the document type. First, provide a plurality of document type distributions, each document type distribution describes layout characteristics of an independent document type and may include a plurality of data points. Each document type distribution includes data derived from at least one basis document signature which may include data defining pixels of a low-resolution image of the independent basis document resolved to between 1 and 75 dots per inch or may include document segmentation data derived from the independent basis document. Next provide a new electronic document. Then create new document signature from the new electronic document. Next, distances between the new document signature and each of the plurality of document type distributions are calculated using an algorithm based on a Bayesian framework for a Gaussian distribution. The distances calculated may be Euclidean distances or may be Mahalanobis distances. Additionally, calculating the distances may include weighting the value given each of a plurality of data points in the document signatures based on the usefulness of each of the plurality of data points in distinguishing between the document signatures. Next, select at least one candidate document type for the new electronic document from among the independent document types described by the plurality of document type distributions. The selection of the at least one candidate document type may include selecting a preselected fixed number of the independent document types or may include selecting the independent document types described by those of the plurality of document type distributions having calculated distances that are within a preselected threshold distance of the smallest of the distances calculated. In addition, the invention provides for a program storage medium readable by computer, tangibly embodying a program of instructions executable by the computer to perform the method steps described above.

EP 1 096 391 A3

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 00 11 8884

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.7)
Y	ESPOSITO F ET AL: "An experimental page layout recognition system for office document automatic classification: an integrated approach for inductive generalization" PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION. ATLANTIC CITY, JUNE 16 - 21, 1990. CONFERENCE A : COMPUTER VISION AND CONFERENCE B : PATTERN RECOGNITION SYSTEMS AND APPLICATIONS, LOS ALAMITOS, IEEE COMP. SOC. PRESS, US, vol. VOL. 1 CONF. 10, 16 June 1990 (1990-06-16), pages 557-562, XP010020366 ISBN: 0-8186-2062-5 * the whole document * -----	1-22	G06F17/30
Y	MILLNS I ET AL: "An integrity constraint for database systems containing embedded neural networks" DATABASE AND EXPERT SYSTEMS APPLICATIONS, 1998. PROCEEDINGS. NINTH INTERNATIONAL WORKSHOP ON VIENNA, AUSTRIA 26-28 AUG. 1998, LOS ALAMITOS, CA, USA,IEEE COMPUT. SOC, US, 26 August 1998 (1998-08-26), pages 56-61, XP010296713 ISBN: 0-8186-8353-8 * the whole document * ----- -/--	1-22	TECHNICAL FIELDS SEARCHED (Int.Cl.7) G06F

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
Munich	2 April 2004	Korkuzas, V

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 00 11 8884

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.7)
A	ESPOSITO F ET AL: "A knowledge-based approach to the layout analysis" DOCUMENT ANALYSIS AND RECOGNITION, 1995., PROCEEDINGS OF THE THIRD INTERNATIONAL CONFERENCE ON MONTREAL, QUE., CANADA 14-16 AUG. 1995, LOS ALAMITOS, CA, USA,IEEE COMPUT. SOC, US, 14 August 1995 (1995-08-14), pages 466-471, XP010230978 ISBN: 0-8186-7128-9 * the whole document *	1-22	
A	BADAL D Z ED - INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS: "Neural network recognition of human face images stored in the database" PROCEEDINGS OF THE INTERNATIONAL PHOENIX CONFERENCE ON COMPUTERS AND COMMUNICATIONS. TEMPE, MAR. 23 - 26, 1993, NEW YORK, IEEE, US, vol. CONF. 12, 23 March 1993 (1993-03-23), pages 552-558, XP010110100 ISBN: 0-7803-0922-7 * the whole document *	1-22	TECHNICAL FIELDS SEARCHED (Int.Cl.7)

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
Munich	2 April 2004	Korkuzas, V

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)